

As pointed out by Applicants' representative in the telephonic interview, the Sun reference does not suggest "encoding . . . biological molecules into a data structure of initial character strings." In fact, nothing in the Sun reference suggests entering or manipulating such character strings. As explained, the Sun reference describes an analytical model DNA recombination. Specifically, it provides a model of DNA fragmentation and assembly, including probability of having mutations in fragments being recombined, probability of recombination, etc. The parameters used in the model include distance of interest on the genome, fragment length, fractional overlap, relative locations of mutations, etc. Sun's disclosed parameters are all that are required for his model. The disclosed parameters do not include a character string encoded from biological molecules.

The ultimate goals of the Sun model include identifying optimal fragment sizes, distances between mutations, etc. These goals can be realized without using character strings corresponding to encoded biological molecules. And it is respectfully submitted that nothing in the Sun reference suggests use or manipulation of such character strings. It is believed therefore that all claims are patentable over the Sun reference.

The Examiner has pointed out that Sun mentions "concatenation of random fragments" at page 81. The relevant sentence reads as follows: "Then any full length reassembled molecule can be regarded as concatenations of random fragments." This merely states a principle of DNA reassembly. It is a premise applied by Sun in developing his models. It is not a suggestion that one should encode molecules as character strings.

Applicants believe that all pending claims are allowable and respectfully requests a Notice of Allowance for this application from the Examiner. Should the Examiner believe that a telephone conference would expedite the prosecution of this application, the undersigned can be reached at the telephone number set out below.

Respectfully submitted,

BEYER, WEAVER & THOMAS, LLP

A handwritten signature in black ink, appearing to read "Jeffrey K. Weaver", with a long, sweeping horizontal line extending to the right.

Jeffrey K. Weaver  
Reg. No. 31,314

P.O. Box 778  
Berkeley, CA 94704-0778  
(510) 843-6200

**APPENDIX 1**  
**MARKED UP VERSION OF AMENDED CLAIMS**

1. (Twice Amended)            A method of **identifying for production** [producing] molecules represented by concatenated strings, said method comprising:

i) encoding two or more related biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;

ii) selecting at least two substrings from said initial character strings;

iii) concatenating said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adding the product strings to a data structure to populate a data structure of product strings;

v) [optionally repeating steps (i) or (ii) through (iv);) using one or more of said product strings as an initial string in the collection of initial character strings, and;] **determining whether the product strings have sequence identities of greater than 30% with at least one initial character string; and**

vi) **selecting for production** [making] one or more product biological molecules corresponding to one or more of the product strings **determined to have greater than 30% sequence identity with at least one initial string.**

17. (Twice Amended)            A computer program product on a computer readable media comprising computer code that:

i) encodes two or more biological molecules into initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about ten subunits;

ii) selects at least two initial substrings from said character strings;

iii) concatenates said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adds the product strings to a data structure to populate a data structure of product strings;

v) [optionally repeats steps (i) or (ii) through (iv) using one or more of said product strings as an initial string in the collection of initial character strings; and,] **determines whether**

**the product strings have sequence identities of greater than 30% with at least one initial character string; and**

vi) **selects for** [directs the] production of one or more product biological molecules corresponding to **one or more of** the product strings **determined to have greater than 30% sequence identity with at least one initial string.**

**APPENDIX 2**  
**PENDING CLAIMS**

1. (Amended) A method of identifying for production molecules represented by concatenated strings, said method comprising:

i) encoding two or more related biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;

ii) selecting at least two substrings from said initial character strings;

iii) concatenating said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adding the product strings to a data structure to populate a data structure of product strings; and

v) determining whether the product strings have sequence identities of greater than 30% with at least one initial character string; and

vi) selecting for production one or more product biological molecules corresponding to one or more of the product strings determined to have greater than 30% sequence identity with at least one initial string.

2 (Amended). The method of claim 1, wherein said encoding comprises encoding two or more nucleic acid sequences into said character strings.

3 (Amended). The method of claim 2, wherein said two or more nucleic acid sequences comprise a nucleic acid sequence encoding a naturally occurring protein.

4 (Amended). The method of claim 1, wherein said encoding comprises encoding two or more amino acid sequences into said character strings.

5 (Amended). The method of claim 4, wherein said one or more amino acid sequences comprise an amino acid sequence encoding a naturally occurring protein.

6. The method of claim 1, wherein said biological molecules have at least 30% sequence identity with each other.

7. The method of claim 1, wherein said selecting comprises selecting substrings such that the ends of said substrings occur in string regions of about 3 to about 20 characters that have higher sequence identity with the corresponding region of another of said initial character strings than the overall sequence identity between the same two strings.

8. The method of claim 1, wherein said selecting comprises selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

9. The method of claim 1, wherein said selecting and concatenating comprises concatenating substrings from two different initial strings such that the concatenation occurs in a region of about three to about twenty characters having higher sequence identity between said two different initial strings than the overall sequence identity between said two different initial strings.

10 (Amended). The method of claim 1, wherein said selecting comprises aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

11 Canceled.

12 (Amended). The method of claim 1, wherein said method further comprises randomly altering one or more characters of said initial or product character strings.

13. The method of claim 12, wherein said method further comprises randomly selecting and altering one or more occurrences of a particular preselected character in said character strings.

14 (Amended). The method of claim 1, wherein said encoding, selecting, or concatenating is performed on an internet site.

15 (Amended). The method of claim 1, wherein said encoding, selecting, or concatenating is performed on a server.

16 (Amended). The method of claim 1, wherein said encoding, selecting, or concatenating is performed on a client linked to a network.

17. (Amended) A computer program product on a computer readable media comprising computer code that:

- i) encodes two or more biological molecules into initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about ten subunits;
- ii) selects at least two initial substrings from said character strings;
- iii) concatenates said substrings to form one or more product strings about the same length as one or more of the initial character strings;
- iv) adds the product strings to a data structure to populate a data structure of product strings; and
- v) determines whether the product strings have sequence identities of greater than 30% with at least one initial character string; and
- vi) selects for production of one or more product biological molecules corresponding to one or more of the product strings determined to have greater than 30% sequence identity with at least one initial string.

18 (Amended). The computer program product of claim 17, wherein said two or more biological molecules are nucleic acid sequences.

19 (Amended). The computer program product of claim 17, wherein said two or more biological molecules are nucleic acid sequences of naturally occurring proteins.

20 (Amended). The computer program product of claim 17, wherein said two or more biological molecules are amino acid sequences.

21 (Amended). The computer program product of claim 17, wherein said biological molecules have at least 30% sequence identity.

22 (Amended). The computer program product of claim 17, wherein said computer code selects substrings such that the ends of said substrings occur in string regions of about three to about twenty characters that have higher sequence identity with a corresponding region of another of said initial character strings than the overall sequence identity between the two substrings.

23 (Amended). The computer program product of claim 17, wherein said computer code selects substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

24 (Amended). The computer program product of claim 17, wherein said computer code selects and concatenates substrings from two different initial strings such that the concatenation occurs in a region of about three to about twenty characters having higher sequence identity between said two different initial strings than the overall sequence identity between said two different initial strings.

25 (Amended). The computer program product of claim 17, wherein the computer code selects substrings by aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the character strings, and selecting a character that is a member of an aligned pair for the end of one substring.

26 Canceled.

27 (Amended). The computer program product of claim 17, wherein said computer code additionally randomly alters one or more characters of said character strings.

28 (Amended). The computer program product of claim 27, wherein said computer code additionally randomly selects and alters one or more occurrences of a particular preselected character in said character strings.

29 (Amended). The computer program product of claim 17, wherein said computer code is stored on media selected from the group consisting of magnetic media, optical media, optomagnetic media.

30 (Amended). The computer program product of claim 17, wherein said computer code is in dynamic or static memory of a computer.

45 (New). The method of claim 1, wherein the initial character strings of (i) are related.

46 (New). The method of claim 1, further comprising physically screening the molecule(s) represented by the product strings for one or more desired properties.

47 (New). The method of claim 1, further comprising determining a computationally predicted property for molecules represented by the product strings.

48 (New). The method of claim 1, wherein the molecules represented by the product strings are made in parallel in an array of vessels.

49 (New). The method of claim 1, wherein the molecules represented by the product strings are made by assembly of oligonucleotides.

50 (New). The method of claim 1, further comprising testing members of the data structure of product strings for a particular property and determining an optimal combination of sequences using multi-variate analysis.

51 (New). The computer program product of claim 17, wherein the initial character strings of (i) are related.

52 (New). The computer program product of claim 17, wherein the code instructs physical screening of the molecule(s) represented by the product strings for one or more desired properties.

53 (New). The computer program product of claim 17, wherein the code instructs determination of a computationally predicted property for molecules represented by the product strings.

54 (New). The computer program product of claim 17, wherein the molecules represented by the product strings are made in parallel in an array of vessels.

55 (New). The computer program product of claim 17, wherein the molecules represented by the product strings are made by assembly of oligonucleotides.

56 (New). The computer program product of claim 17, wherein the code tests members of the data structure of product strings for a particular property and determines an optimal combination of initial sequences using multi-variate analysis.